

A NONSTATIONARY NONPARAMETRIC BAYESIAN APPROACH TO DYNAMICALLY MODELING EFFECTIVE CONNECTIVITY IN FUNCTIONAL MAGNETIC RESONANCE IMAGING EXPERIMENTS

BY SOURABH BHATTACHARYA AND RANJAN MAITRA¹

Indian Statistical Institute and Iowa State University

Effective connectivity analysis provides an understanding of the functional organization of the brain by studying how activated regions influence one other. We propose a nonparametric Bayesian approach to model effective connectivity assuming a dynamic nonstationary neuronal system. Our approach uses the Dirichlet process to specify an appropriate (most plausible according to our prior beliefs) dynamic model as the “expectation” of a set of plausible models upon which we assign a probability distribution. This addresses model uncertainty associated with dynamic effective connectivity. We derive a Gibbs sampling approach to sample from the joint (and marginal) posterior distributions of the unknowns. Results on simulation experiments demonstrate our model to be flexible and a better candidate in many situations. We also used our approach to analyzing functional Magnetic Resonance Imaging (fMRI) data on a Stroop task: our analysis provided new insight into the mechanism by which an individual brain distinguishes and learns about shapes of objects.

1. Introduction. Functional magnetic resonance imaging (fMRI) is a non-invasive technique for detecting regions in the brain that are activated by the application of a stimulus or the performance of a task. Although important neuronal activities are responsible for such activation, these are very subtle and can not be detected directly. Instead, local changes during neuronal activity in the flow, volume, oxygen level and other characteristics of blood, called the blood oxygen level dependent (BOLD) response, form a proxy. Much research in fMRI has focused on identifying regions of cerebral activation in response to the activity of interest. There is, however, growing interest

Received October 2010; revised February 2011.

¹Supported in part by NSF CAREER Grant DMS-04-37555 and by the National Institutes of Health (NIH) Award DC-0006740.

Key words and phrases. Attentional control network, Bayesian analysis, Dirichlet process, effective connectivity analysis, fMRI, Gibbs sampling, temporal correlation.

This is an electronic reprint of the original article published by the Institute of Mathematical Statistics in *The Annals of Applied Statistics*, 2011, Vol. 5, No. 2B, 1183–1206. This reprint differs from the original in pagination and typographic detail.

in obtaining better understanding of the interactions between different brain regions during the operation of the BOLD response. The study of how one neuronal system interacts with another is called effective connectivity analysis [Friston (1994); Nyberg and McIntosh (2001)]. We illustrate this in the context of obtaining greater insight into how an individual brain performs a Stroop task, which is also the main application studied in this paper.

1.1. *Investigating the attentional control network in a Stroop task.* The human brain’s information processing capability is limited, so it sifts out irrelevant details from task-relevant information using the cognitive function called *attention*. Specifically, task-relevant information is filtered out either because of intrinsic properties of the stimulus (bottom-up selection) or independently (top-down selection) [Frith (2001)]. The brain’s preference for task-related information in top-down selection requires coordination of neural activity via an Attentional Control Network (ACN) which has systems to process task-relevant and irrelevant information and also a “higher-order executive control system” to modulate the frequency of neuronal firings in each [Banach et al. (2000)]. Thus, the higher-order system can execute top-down selection by increasing neuronal activity in the task-relevant processing system while suppressing it in its task-irrelevant counterpart. Many studies have empirically found the dorsal lateral prefrontal cortex (DLPFC) to be the main source of attentional control, while the task-relevant and irrelevant processing sites depend on whether the stimulus is visual, auditory or in some other form.

Jaensch (1929) and Stroop (1935) discovered that the brain is quicker at reading named color words (e.g., blue, yellow, green, etc.) when they are in the concordant color than if they are in a discordant color. Tasks structured along these lines are now called Stroop tasks. A much-studied two-phase experiment [Milham et al. (2002, 2003); Ho, Ombao and Shumway (2003, 2005); Milham, Banich and Barad (2003); Bhattacharya, Ho and Purkayastha (2006)] designed around such a task provided the dataset for our investigation. In the first phase, a subject was trained to associate each of three unfamiliar shapes with a unique color word (“Blue,” “Yellow” and “Green”) with 100% accuracy. The second (testing) phase involved alternating six times between blocks of eighteen *interference* and eighteen *neutral* trials. The neutral trial consisted of printing the shape in a neutral color (white). The interference trial involved presenting the subject with one of the learned shapes, but printed in a color different from that learned to be represented by that shape in the learning phase. The subject’s task was to subvocally name the shape’s color as trained in the learning phase, ignoring the color presented in the testing phase. Each neutral or interference trial consisted of a 0.3-s fixation cross, a 1.2-s stimulus presentation stage and a 0.5-s waiting state till the next trial. fMRI images were acquired and

processed to obtain three activated regions, whose averaged post-processed time series are what we analyze further to investigate attentional control. These three regions—also denoted as Regions 1, 2 and 3 in this paper—were the lingual gyrus (LG), the middle occipital gyrus (MOG) and the DLPFC, and were chosen as representatives of task-irrelevant, task-relevant and executive-control systems, respectively. The LG is a visual area for processing color information [Corbetta et al. (1991)], which in our context is task-irrelevant [Kelley et al. (1998)]. The MOG is another visual area but processes shape information, which is the task-related information (form of the shape) in the experiment. We refer to Bhattacharya, Ho and Purkayastha (2006) for further details on data collection and post-processing, noting here that, as in that and other preceding papers, the objective is to investigate and to understand the working of the ACN mechanism in performing a Stroop task.

1.2. Background and related work. Structural equation modeling [McIntosh and Gonzalez-Lima (1994); Kirk et al. (2005); Penny et al. (2004)] and time-varying parameter regression [Büchel and Friston (1998)] are two early approaches that have been used to determine effective connectivity. In general, both approaches ignore dynamic modeling of the observed system, even though the latter accounts for temporal correlation in the analysis. There is, however, strong empirical evidence [Aertsen and Preißl (1991); Friston (1994); McIntosh and Gonzalez-Lima (1994); Büchel and Friston (1998); McIntosh (2000)] that effective connectivity is dynamic in nature, which means that the time-invariant model assumed by both approaches may not be appropriate. Ho, Ombao and Shumway (2005) overcame some of these limitations by modeling the data using a state-space approach, but did not account for the time-varying nature of the effective connectivity parameters.

An initial attempt at explicitly incorporating the time-varying nature of effective connectivity in addition to dynamic modeling of neuronal systems was by Bhattacharya, Ho and Purkayastha (2006), who adopted a Bayesian approach to inference and developed and illustrated their methodology with specific regard to the ACN mechanism of the LG, MOG and DLPFC regions in conducting the Stroop task outlined above. We summarize their model—framing it within the context of more recent literature in dynamic modeling of effective connectivity—and discuss their findings and some limitations next. In doing so, we also introduce the setup followed throughout this paper.

1.2.1. Bayesian modeling of dynamic effective connectivity. Let $y_i(t)$ be the observed fMRI signal (or the measured BOLD response) corresponding to the i th region at time t , $i = 1, 2, \dots, R$, $t = 1, 2, \dots, T$. Specifically, $y_i(t)$ is some voxel-wise summary (e.g., regional average) of the corresponding detrended time series in the i th region. Following Bhattacharya, Ho and

Purkayastha (2006), let $x_i(t)$ be the modeled BOLD response [as opposed to the measured BOLD response, $y_i(t)$], that is, the stimulus $s(t)$ convolved with the hemodynamic response function (HRF) $h_i(t)$ for the i th region and time point t . In this paper, $h_i(t)$ is assumed to be the very widely-used standard HRF model of Glover (1999) which differences two gamma functions and has some very appealing properties vis-a-vis other HRFs [Lu et al. (2006, 2007)]. Then the model for the observed fMRI signal can be hierarchically specified as

$$(1) \quad y_i(t) = \alpha_i + x_i(t)\beta_i(t) + \varepsilon_i(t),$$

where α_i and $\beta_i(t)$ are the baseline trend and activation coefficients for the i th region, the latter at time t . The errors $\varepsilon_i(t)$'s are all assumed to be independent $N(0, \sigma_i^2)$, following Worsley et al. (2002). From Bhattacharya, Ho and Purkayastha [(2006), page 797], we assume that $x_i(\cdot) = x(\cdot)$ for $i = 1, \dots, R$, that is, we use the same HRF $h_i(\cdot) = h(\cdot)$ for each of the R regions. Note that, as argued in that paper, this homogeneous assumption on the $x(\cdot)$ is inconsequential because it is compensated by the $\beta_i(t)$ that are associated with $x(t)$, and allowed to be inhomogeneous with respect to the different regions. Also, following Bhattacharya, Ho and Purkayastha [(2006), page 799], we assume that $\sigma_i^2 = \sigma_\varepsilon^2$; $i = 1, \dots, R$. Actually, (1) is a generalization of a very standard model used extensively in the literature—see, for example, Lindquist [(2008), equation (9)] or Henson and Friston [(2007), page 179, equation (14.1)], who use the same model but with a constant time-invariant $\beta(t) \equiv \beta$. (Indeed, as very helpfully pointed out by a reviewer, this last specification is also the general linear model commonly used to analyze fMRI data voxel-wise, such as in statistical parametric mapping and related conventional whole brain activation studies.) Our specific generalization incorporates time-varying $\beta(t)$ and follows Ho, Ombao and Shumway (2005), Bhattacharya, Ho and Purkayastha (2006) or Harrison, Stephan and Friston [(2007), cf. page 516, equation 38.18]—note, however, that the latter model $\beta(t)$ as a random walk [see equation 38.19, page 516, of Harrison, Stephan and Friston (2007)]. We prefer allowing for time-varying activation $\beta_i(t)$ in order to address the “learning” effect often reported in fMRI studies whereby strong activation in the initial stages of the experiment dissipates over time [Gössl, Auer and Fahrmeir (2001); Milham et al. (2002, 2003); Milham, Banich and Barad (2003)]. Further modeling specifies the activation coefficient in the i th region at the t th time-point in terms of the noise-free BOLD signal in the other regions at the previous time-point. Thus,

$$(2) \quad \beta_i(t) = x(t-1) \left[\sum_{\ell=1}^R \gamma_{i\ell}(t) \beta_\ell(t-1) \right] + \omega_i(t),$$

$$t = 2, \dots, T; i = 1, 2, \dots, R,$$

where $\omega_i(t)$ are independent $N(0, \sigma_\omega^2)$ -distributed errors and $\gamma_{ij}(t)$ is the influence of the j th region on the i th region at time t . Under (2), functionally specified cerebral areas are not constrained to act independently but can interact with other regions. Our objective is to make inferences on $\gamma_{ij}(t)$ in order to understand the functional circuitry in the brain as it processes a certain (in this paper, Stroop) task.

Equations (1) and (2) together specify one of many Vector Autoregressive (VAR) models proposed by several authors [Harrison, Penny and Friston (2003); Goebel et al. (2003); Rykhlevskaia, Fabiani and Gratton (2006); Sato et al. (2007); Thompson and Siegle (2009); Patriota, Sato and Achic (2010)]. To see this, note that for $i = 1, \dots, R$, $\beta_i(t-1)$ depends linearly upon $y_i(t-1)$. Hence, substituting this in (2) yields $\beta_i(t) = g_i(y_1(t-1), y_2(t-1), \dots, y_R(t-1))$, for known functions g_i , which are linear in $y_1(t-1), y_2(t-1), \dots, y_R(t-1)$. Then substituting $\beta_i(t)$ in (1), we see that for each $i = 1, \dots, R$, $y_i(t)$ is a linear function of $y_1(t-1), y_2(t-1), \dots, y_R(t-1)$. Hence, the vector $\mathbf{y}(t) = (y_1(t), \dots, y_R(t))'$ is a linear function of the vector $\mathbf{y}(t-1) = (y_1(t-1), \dots, y_R(t-1))'$. As a result, our model is a first order VAR model from the viewpoint of the responses. It is of first order since $\mathbf{y}(t)$ depends upon $\mathbf{y}(t-1)$, given $\mathbf{y}(1), \dots, \mathbf{y}(t-1)$. Moreover, (2) shows that the activation coefficients $\beta_i(t)$ are modeled as first order VAR; that is, the R -component vector $(\beta_1(t), \dots, \beta_R(t))'$ depends linearly upon $(\beta_1(t-1), \dots, \beta_R(t-1))'$.

VAR models provide an alternative or a substantial generalization [Friston (2009)] to the Dynamic Causal Modeling (DCM) approach proposed by Friston, Harrison and Penny (2003), at least in continuous-time, to model the change of the neuronal state vector over time, using stochastic differential equations. In DCM, the observed BOLD signal is modeled as $y_i(t) = r_i(t) + \beta z_i(t) + \varepsilon_i(t)$, where $z_i(t)$ denotes nuisance effects, and $r_i(t)$ is a modeled BOLD response obtained by first using a bilinear differential (neural state) equation, parametrized in terms of effective connectivity parameters and involving $s(t)$, then subsequently using a “balloon model” transformation [Buxton, Wong and Frank (1998) or extensions Friston et al. (2000); Stephan et al. (2007)] to the solution of the bilinear differential equation. DCM thus uses both $r_i(t)$ as well as the nuisance effects $z_i(t)$ to model the observed BOLD response, with $r_i(t)$ playing the same role as our $x_i(t)$ with the exception that the latter is obtained using the more widely-used Glover (1999) HRF model. Further, DCM assumes a deterministic relationship between the different brain regions unlike (2) which allows for noisy dynamics [Bhattacharya, Ho and Purkayastha (2006)].

Thompson and Siegle (2009) contend that VAR models have gained popularity in recent years because “the direction and valence of effective connectivity relationships do not need to be pre-specified.” As such, these models have provided a useful framework for effective connectivity analysis.

Bhattacharya, Ho and Purkayastha (2006) proposed a symmetric random walk model for $\gamma_{ij}(t)$:

$$(3) \quad \gamma_{ij}(t) = \gamma_{ij}(t-1) + \delta_{ij}(t) \quad \text{for } i, j = 1, 2, \dots, R; t = 2, 3, \dots, T,$$

where δ_{ij} are independent $N(0, \sigma_\delta^2)$ -distributed errors. In this paper we use \mathcal{M}_{RW} to refer to the model specified by (1), (2) and (3). The effective connectivity parameters $\gamma_{ij}(t)$; $(i, j) = 1, \dots, R$, also form a VAR model of the first order. To see this, let $\mathbf{\Gamma}(t) = (\gamma_{ij}(t); i, j = 1, \dots, R)'$. Then it follows that $\mathbf{\Gamma}(t) = \mathbf{I}\mathbf{\Gamma}(t-1) + \mathbf{\delta}(t)$, where \mathbf{I} is the $R \times R$ -order identity matrix and $\mathbf{\delta}(t) = (\delta_{ij}(t); i, j = 1, \dots, R)'$, indicating that $\gamma_{ij}(t)$'s are within the framework of a VAR model.

Bhattacharya, Ho and Purkayastha (2006) specified prior distributions on the parameters and hyperparameters of this model and used Gibbs sampling to learn the posterior distributions of the unknowns. We refer to that paper for details and for results on simulation experiments using \mathcal{M}_{RW} , noting here only that their Bayesian-derived inference supported ACN theory and, more importantly, the notion that effective connectivity is indeed dynamic in the network. Further, they found that the restricted model with $\gamma_{31}(t) = \gamma_{32}(t) \equiv 0 \forall t$ was the best-performer, implying no direct feedback from the two sites of control (LG and MOG) to the source (DLPFC). Interestingly, however, and perhaps surprisingly, their estimated $\gamma_{ij}(t)$'s (see Figure 6 in their paper) had very little relationship with the nature of the BOLD response (see Figure 1, bottom panel, in that paper). This is surprising because from (1), we have $\beta_i(t) = (y_i(t) - \alpha_i - \varepsilon_i(t))/x(t)$, and similarly for $\beta_i(t-1)$, which when substituted on the right-hand side of (2) makes it independent of $x(\cdot)$. This means that the effective connectivity parameters $\gamma_{i\ell}(t)$ depend upon $\beta_i(t)$, the left-hand side of (2). Since $\beta_i(t)$ is a function of $x(t)$, it is reasonable to expect $\gamma_{i\ell}(t)$'s to depend upon $x(t)$, but such a relationship was not found in Bhattacharya, Ho and Purkayastha (2006). This perplexing finding led us to first investigate robustness of \mathcal{M}_{RW} to even slight misspecifications.

1.2.2. Robustness of the random walk model. We tested the effect of a slight departure from \mathcal{M}_{RW} by simulating, instead of from (3), from the following stationary autoregressive model:

$$(4) \quad \gamma_{ij}(t) = 0.999\gamma_{ij}(t-1) + \delta_{ij}(t) \quad \text{for } i, j = 1, 2, \dots, R; t = 2, 3, \dots, T.$$

We call this slightly modified model $\mathcal{M}_{\text{RW}'}$. Here, $T = 285$ and $R = 3$ to match the details of the dataset of Section 1.1. We fit \mathcal{M}_{RW} to data simulated from $\mathcal{M}_{\text{RW}'}$. Figure 1 displays the estimated posterior distributions of $\gamma_{ij}(t)$. The marginal posterior distribution of each $\gamma_{ij}(t)$ is represented here by eight quantiles, each containing 12.5% of the distribution: increased

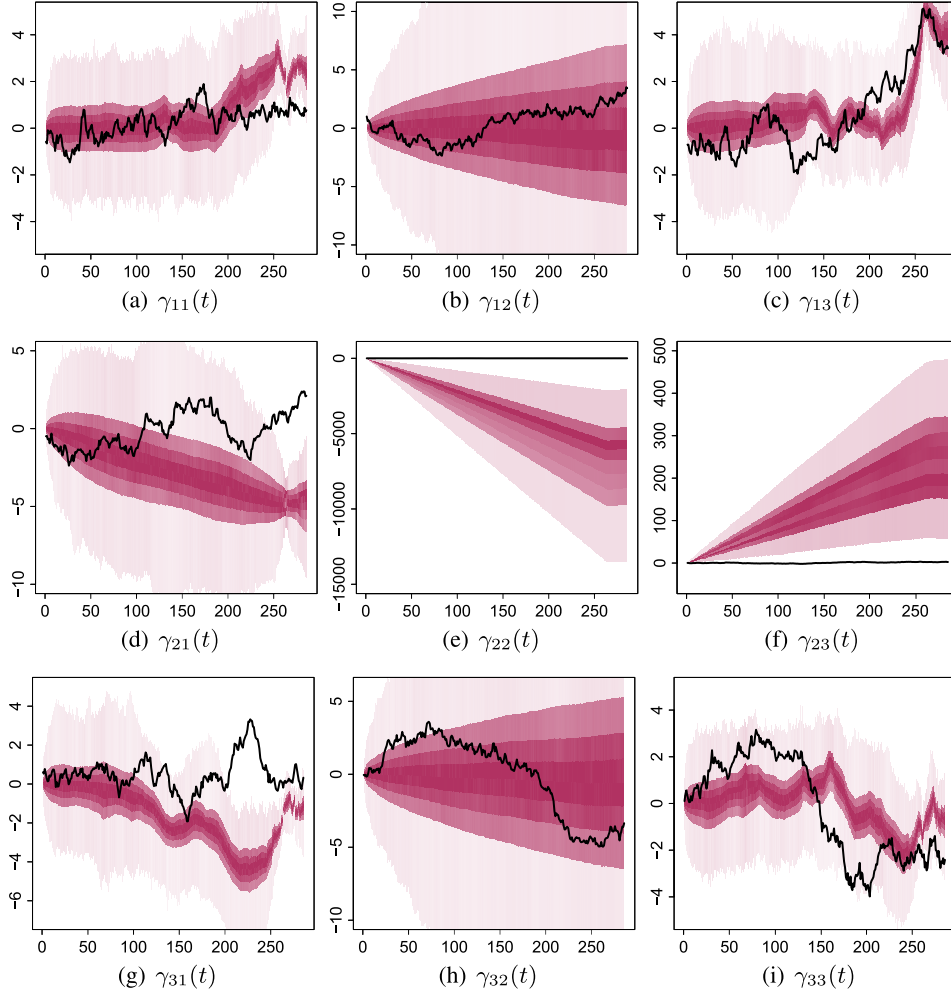


FIG. 1. *Posterior densities of $\gamma_{ij}(t); t = 1, \dots, T; i, j = 1, 2, 3$, under model \mathcal{M}_{RW} on data simulated under model $\mathcal{M}_{\text{RW}'}$. The opacity of shading in each region is proportional to the area under the density in that region. The solid line stands for the true values of $\gamma_{ij}(t)$.*

opacity in shading denotes denser regions. Solid lines represent true values. As seen, many parts of the posterior distribution have very little coverage of the true effective connectivity parameters: this finding is also supported by Table 1 which provides the proportion of true values included in the 95% highest posterior density (HPD) credible intervals [Berger (1985)] (these are the shortest intervals with posterior probability 0.95). Thus, performance degrades substantially even though $\mathcal{M}_{\text{RW}'}$ is not all that different from \mathcal{M}_{RW} . Hence, modeling the process by a random walk may be too restrictive and thus a better approach may be needed. We do so in this paper by embed-

TABLE 1
Proportion of true $\gamma_{ij}(t)$ included in the 95% posterior credible intervals obtained using model \mathcal{M}_{RW} on data simulated using \mathcal{M}_{RW}

γ_{11}	γ_{12}	γ_{13}	γ_{21}	γ_{22}	γ_{23}	γ_{31}	γ_{32}	γ_{33}
0.99	0.99	0.91	1.0	0	0.05	0.05	1.0	0.60

ding an (asymptotically) stationary first order autoregressive AR(1) model in a larger class of models. Formally, we employ a Bayesian nonparametric framework using a Dirichlet Process (DP) prior whose base distribution is assumed to be that implied by an AR(1) model. The intuition behind this modeling style is that although one might expect the actual process to be stationary, the assumption might be too simplistic, and it is more logical to think of the stationary model as an “expected model,” thus allowing for nonstationarity (quantified by the DP prior) in the actual model. Theoretical issues related to the construction of DP-based nonstationary processes are discussed in Section 2.1. In Section 2.2 we introduce our new modeling ideas using the developments in Section 2.1. The efficacy of the new model is compared with its competitors on some simulated datasets in Section 3. The new approach is applied in Section 4 to the dataset introduced in Section 1.1 to investigate effective connectivity between the LG, MOR and DLPFC regions. We conclude in Section 5 with some discussion. Additional derivations and further details on experiments and data analyses are provided in the supplement [Bhattacharya and Maitra (2011)], whose sections, figures and tables have the prefix “S-” when referred to in this paper.

2. Modeling and methodology.

2.1. A nonstationary Dirichlet process model for time series observations.

A random probability measure \mathbf{G} on the probability space $(\mathbf{\Gamma}, \mathcal{B}_{\mathbf{\Gamma}})$ sampled from the Dirichlet Process (DP) denoted by $\text{DP}(\tau \mathbf{G}_0)$, and with known distribution \mathbf{G}_0 and precision parameter τ , can be represented almost surely, using the constructive method provided in Sethuraman (1994), as

$$(5) \quad \mathbf{G} \equiv \sum_{k=1}^{\infty} p_k \delta_{\gamma_k^*},$$

where $p_1 = b_1$ and $p_k = b_k \prod_{\ell=1}^{k-1} (1 - b_{\ell})$, $k = 2, 3, \dots$, with b_k ’s being independent, identically distributed (henceforth i.i.d.) $\beta(1, \tau)$ random variables. The values γ_k^* are i.i.d. realizations from \mathbf{G}_0 , for $k = 1, 2, \dots$, and are also independent of $\{b_1, b_2, \dots\}$. Note that (5) implies that \mathbf{G} is discrete with probability one, and has expectation \mathbf{G}_0 . DPs thus provide ways to place priors on probability measures.

The dependent Dirichlet process (DDP) is an extension of the DP in the sense that it allows for a prior distribution to be specified on a set of random probability measures, rather than on a single random probability measure. In other words, the realizations γ_k^* can be extended to accommodate an entire time-series domain \mathcal{T} , such that $\mathbf{\Gamma}_{k,\mathcal{T}}^* = \{\gamma_{kt}^*; t \in \mathcal{T}\}$. Following (5), the random process thus constructed can be represented as

$$(6) \quad \mathbf{G}^{(\mathcal{T})} \equiv \sum_{k=1}^{\infty} p_k \delta_{\mathbf{\Gamma}_{k,\mathcal{T}}^*}$$

with form similar to that used for spatial DP models [see Gelfand, Kottas and MacEachern (2005)]. Note that $\mathbf{\Gamma}_{k,\mathcal{T}}^*$ in (6) are realizations of some stochastic process $\mathbf{\Gamma}_{\mathcal{T}} = \{\gamma_t; t \in \mathcal{T}\}$, with distribution $\mathbf{G}_0^{(\mathcal{T})}$ for $k = 1, 2, \dots$. Hence, Kolmogorov's consistency holds for $\mathbf{\Gamma}_{\mathcal{T}}$. That is, finite dimensional joint distributions $\{\gamma_t; t \in \mathbf{t}_T\}$, for ordered time-points $\mathbf{t}_T = \{t_1, \dots, t_T\}$, can be obtained from all finite but higher-dimensional joint distributions $\{\gamma_t; t \in \mathbf{t}_T^* \cup \mathbf{t}_T\}$ (here \mathbf{t}_T^* is a finite set) specified by the process, by marginalizing over $\{\gamma_t; t \in \mathbf{t}_T^*\}$. Since (6) shows that $\mathbf{G}^{(\mathcal{T})}$ is specified completely by the process $\mathbf{\Gamma}_{\mathcal{T}}$ and $\{p_k; k = 1, 2, \dots\}$, and since the latter are independent of t , it follows that Kolmogorov's consistency holds for $\mathbf{G}^{(\mathcal{T})}$, providing a formal setup of a stochastic process of random distributions. In particular, for any $t \in \mathcal{T}$, $\mathbf{G}^{(\{t\})} \sim \text{DP}(\tau \mathbf{G}_0^{(\{t\})})$ [and admits the representation $\mathbf{G}^{(\{t\})} \equiv \sum_{k=1}^{\infty} p_k \delta_{\gamma_{kt}^*}$]. The collection of random measures $\mathbf{G}^{(\mathcal{T})}$ is said to follow the DDP [see, e.g., MacEachern (2000); De Iorio et al. (2004); Gelfand, Kottas and MacEachern (2005)].

The process $\mathbf{\Gamma}_{\mathcal{T}}$ may be a time series that is stationary or—as adopted in our application and more realistically—asymptotically so. Indeed, while asymptotic stationarity is a very slight departure from stationarity, Section 1.2.2 demonstrates that it can have quite a significant impact on inference. It is also important to observe that although the process may be stationary or asymptotically stationary under $\mathbf{G}_0^{(\mathcal{T})}$, the same process when conditioned on $\mathbf{G}^{(\mathcal{T})}$ is not even asymptotically stationary. Specifically,

$$E(\gamma_t | \mathbf{G}^{(\mathcal{T})}) = \sum_{k=1}^{\infty} p_k \gamma_{kt}^*, \quad \text{Var}(\gamma_t | \mathbf{G}^{(\mathcal{T})}) = \sum_{k=1}^{\infty} p_k (\gamma_{kt}^*)^2 - \left(\sum_{k=1}^{\infty} p_k \gamma_{kt}^* \right)^2$$

and

$$\text{Cov}(\gamma_s, \gamma_t | \mathbf{G}^{(\mathcal{T})}) = \sum_{k=1}^{\infty} p_k \gamma_{ks}^* \gamma_{kt}^* - \left(\sum_{k=1}^{\infty} p_k \gamma_{ks}^* \right) \left(\sum_{k=1}^{\infty} p_k \gamma_{kt}^* \right).$$

Thus, $\mathbf{G}^{(\mathcal{T})}$ is nonstationary, although under $\mathbf{G}_0^{(\mathcal{T})}$, $\mathbf{\Gamma}_{\mathcal{T}}$ may have a stationary model so that the mean is constant and the covariance depends upon

time only through the time lag $|t - s|$. Thus, we have defined here a process $\mathbf{G}^{(\mathcal{T})}$ that is centered around a stationary process, but is actually nonstationary. For purposes of applications, we have given (ordered) time-points (t_1, \dots, t_T) , a T -variate distribution $\mathbf{G}^{(T)}$ on the space of all T -variate distributions of $(\gamma_1, \dots, \gamma_T)'$ with mean $\mathbf{G}_0^{(T)}$ being the T -variate distribution implied by a standard time series.

The development of our nonstationary temporal process here technically resembles that of a similar spatial process in Gelfand, Kottas and MacEachern (2005), but differs from the latter in that it is actually embedded in the model for the observed fMRI signals. As a result, the full conditional distributions of $\gamma_{ij}(t)$'s in our model are much more general and complicated than similar derivations following Gelfand, Kottas and MacEachern (2005). Another important difference between our approach and that of Gelfand, Kottas and MacEachern (2005) is that the latter had to introduce a pure error (“nugget”) process to avoid discreteness of the distribution of their spatial data. Such discreteness of the distribution (of our temporal data) is naturally avoided here, however, owing to the embedding approach used in our modeling. Gelfand, Kottas and MacEachern (2005) also rely on the availability of replications of the spatial dataset: our embedding approach obviates this requirement by merely assuming the availability of replicated (unobserved) random processes. We now introduce our dynamic effective connectivity model.

2.2. A Dirichlet process-based dynamic effective connectivity model.

2.2.1. Hierarchical modeling. For $i, j = 1, 2, \dots, R$, define the T -component vectors $\mathbf{\Gamma}_{ij} = (\gamma_{ij}(1), \gamma_{ij}(2), \dots, \gamma_{ij}(T))'$. Further, let $\mathbf{\Gamma}_{ij}$'s be i.i.d. \mathbf{G} , where $\mathbf{G} \sim \text{DP}(\tau \mathbf{G}_0)$, with τ denoting the scale parameter quantifying uncertainty in the base prior distribution \mathbf{G}_0 . Also, assume that under \mathbf{G}_0 , $\gamma_{ij}(1) \sim N(\bar{\gamma}, \sigma_\gamma^2)$ and for $t = 2, \dots, T$, $\gamma_{ij}(t) = \rho \gamma_{ij}(t-1) + \delta_{ij}(t)$, where $|\rho| < 1$ and $\delta_{ij}(t) \sim N(0, \sigma_\delta^2)$ are i.i.d. for $i, j = 1, 2, \dots, R; t = 1, 2, \dots, T$. It follows that under \mathbf{G}_0 , $\mathbf{\Gamma}_{ij} \sim N_T(\bar{\gamma} \boldsymbol{\mu}_T, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}_T = (1, \rho, \rho^2, \dots, \rho^{T-1})'$ and for $s \leq t$, $\boldsymbol{\Sigma}$ has the (s, t) th element

$$(7) \quad \Sigma_{st} = \rho^{s+t-2} \sigma_\gamma^2 + \rho^{t-s} \sigma_\delta^2 \left(\frac{1 - \rho^{2(s-1)}}{1 - \rho^2} \right).$$

Note that with \mathbf{G}_0 as described above, the process is stationary if we choose $\bar{\gamma} = 0$ and $\sigma_\gamma^2 = \sigma_\delta^2 / (1 - \rho^2)$, otherwise the process converges to stationarity for large s . In other words, under \mathbf{G}_0 , $E(\gamma_{ij}(s)) = E(\rho^{s-1} \gamma_{ij}(1) + \sum_{r=0}^{s-2} \rho^r \delta_{ij}(s-r)) = \rho^{s-1} \bar{\gamma}$ which converges to 0 as $s \rightarrow \infty$, while from (7) it follows that, as $s \rightarrow \infty$ with $t - s < \infty$, $\Sigma_{st} \rightarrow \rho^{t-s} \sigma_\delta^2 / (1 - \rho^2)$. The case for $s > t$ is similar. Using the above developments, we specify our dynamic

effective connectivity model hierarchically, by augmenting (1) and (2) with the following model for $\gamma_{ij}(t)$'s:

$$\mathbf{\Gamma}_{ij} \stackrel{\text{i.i.d.}}{\sim} \mathbf{G}^{(T)} \quad \text{for } i, j = 1, 2, \dots, R, \text{ where } \mathbf{G}^{(T)} \sim \text{DP}(\tau \mathbf{G}_0^{(T)}).$$

Distributional assumptions on $\varepsilon_i(t)$'s, $\omega_i(t)$'s and $\delta_{ij}(t)$'s are as in Section 1.2.1. We use \mathcal{M}_{DP} to refer to this model: note also that as $\tau \rightarrow \infty$, our DP-based model converges to the AR(1) model, which we denote using \mathcal{M}_{AR} . We note in closing that the effective connectivity parameters are AR(1), hence VAR, under the expected distribution of \mathcal{M}_{DP} . Of course, they are trivially also so under \mathcal{M}_{AR} . Note, however, that given a realization of a random distribution from the Dirichlet process, such VAR representation does not hold.

2.2.2. Other prior distributions. We specify independent prior distributions on each of $\sigma_\varepsilon^2, \sigma_w^2, \sigma_\delta^2, \rho, \tau, \alpha_i, \beta_i(1)$ and $\gamma_{ij}(1); i, j = 1, 2, \dots, R$. Specifically, α_i 's are assumed to be i.i.d. $N(\mu_i, \sigma_\alpha^2)$ for $i = 1, 2, \dots, R$ and β_i 's are assumed to be i.i.d. $N(\bar{\beta}, \sigma_\beta^2)$, for $i = 1, 2, \dots, R$. Also, $\gamma_{ij}(1)$'s are independently distributed with mean $\bar{\gamma}$ and variance σ_γ^2 , while ρ is uniformly distributed in $(-1, 1)$, $\tau \sim \Gamma(a_\tau, b_\tau)$ and $\sigma_\varepsilon^{-2}, \sigma_w^{-2}$ and σ_δ^{-2} are each i.i.d. $\Gamma(a, b)$ with density having the functional form. Here $\mu_i, \sigma_\alpha^2, \bar{\beta}$ and $\sigma_\beta^2, \bar{\gamma}$ and $\sigma_\gamma^2, a, b, a_\tau$ and b_τ are all hyperparameters. In our examples, we take $a = b = 0$, reflecting our ignorance of the unknown parameter σ_δ^2 . Although the Gamma priors with $a = b = 0$ are improper, they yielded proper posteriors in our case, vindicated by fast convergence of the corresponding marginal chains and resulting right-skewed posterior density estimates, which are expected of proper posteriors having positive support. For (a_τ, b_τ) we first fix the expected value of $\Gamma(a_\tau, b_\tau)$ (given by a_τ/b_τ) to be such that in the full conditional distribution of $\mathbf{\Gamma}_{ij}$, given by (8), the “expected” probability of simulating a new realization from the “prior” base measure approximately equals the probability of selecting realizations of $\mathbf{\Gamma}_{i'j'}$, for some $(i', j') \neq (i, j)$. Hence, if there are R^2 nonzero $\mathbf{\Gamma}_{ij}$ in the model, then setting $a_\tau = c(R^2 - 1)$ and $b_\tau = c$ serves the purpose. The resulting prior distribution has variance equal to its expectation if $c = 1$. To achieve large variance, we set $c = 0.1$; the associated prior worked well in our examples. We also experimented with $c = 0.01$ and $c = 0.001$ and noted that while the case with $c = 0.1$ provided the best results (see Tables S-1 and S-2), inferences related to the posterior distributions of the observed data were fairly robust with respect to different choices of c . Moreover, the results demonstrate that in terms of percentage of inclusion of the true γ_{ij} 's, all inclusion percentages, with the exception of γ_{32} and γ_{33} , were quite robust with respect to c . The results corresponding to $c = 0.01$ and $c = 0.001$ were quite similar, while those corresponding to $c = 0.1$ yielded better performance. Further, other hyper-

parameters were estimated empirically from the data as in Bhattacharya, Ho and Purkayastha (2006) using Berger's (1985) ML-II approach.

2.2.3. Full conditional distributions. The posterior distribution of the parameters are specified by their full conditionals, which are needed for Gibbs sampling. The full conditional distributions of α_i , $\beta_i(t)$, σ_ε^2 and σ_ω^2 are of standard form (see Section S-1.1), while those of the Γ_{ij} 's require some careful derivation. To describe these, note that, on integrating out $\mathbf{G}^{(T)}$, the prior conditional distribution of Γ_{ij} given $\Gamma_{k\ell}$ for $(k, \ell) \neq (i, j)$ follows a Polya urn scheme, and is given by

$$(8) \quad [\Gamma_{ij} \mid \Gamma_{k\ell}; (k, \ell) \neq (i, j)] \sim \frac{\tau \mathbf{G}_0^{(T)} + \sum_{(k, \ell) \neq (i, j)} \delta_{\Gamma_{k\ell}}}{\tau + \#\{(k, \ell) : (k, \ell) \neq (i, j)\}}.$$

The above Polya urn scheme shows that marginalization with respect to \mathbf{G} induces dependence among Γ_{ij} in the form of clusterings, while maintaining the same stationary marginal $\mathbf{G}_0^{(T)}$ for each Γ_{ij} . For Gibbs sampling we need to combine (8) with the rest of the model to obtain the *full conditional* distribution given all the other parameters and the data. We obtain the full conditionals by first defining, for $i, j = 1, 2, \dots, R$, diagonal matrices $\mathbf{A}_{ij} = \sigma_\omega^{-2} \text{diag}\{0, x^2(1)\beta_j^2(1), x^2(1)\beta_j^2(2), \dots, x^2(T-1)\beta_j^2(T-1)\}$, where diag lists the diagonal elements of the relevant matrix. We also define T -variate vectors \mathbf{B}_{ij} for $i, j = 1, 2, \dots, R$ with first element equal to zero. For $t = 2, \dots, T$ the t th element of \mathbf{B}_{ij} is $B_{ij}(t) = \sigma_\omega^{-2}[\beta_i(t)\beta_j(t-1)x(t-1) - \beta_j(t-1)x^2(t-1)\sum_{\ell=1: \ell \neq j}^R \gamma_{i\ell}(t)\beta_\ell(t-1)]$. Further, we note that, thanks to conditional independence, it is only necessary to combine (8) with (2) to obtain the required full conditionals. It follows that

$$(9) \quad [\Gamma_{ij} \mid \dots] \sim q_0^{(ij)} \mathbf{G}_{ij}^{(T)} + \sum_{(k, \ell) \neq (i, j)} q^{(k\ell)} \delta_{\Gamma_{k\ell}},$$

where $\mathbf{G}_{ij}^{(T)}$ is the T -variate normal distribution with mean $(\Sigma^{-1} + \mathbf{A}_{ij})^{-1}(\bar{\gamma}\Sigma^{-1} \times \boldsymbol{\mu} + \mathbf{B}_{ij})$ and variance $(\Sigma^{-1} + \mathbf{A}_{ij})^{-1}$. Also,

$$(10) \quad q_0^{(ij)} = C \frac{\tau}{|\mathbf{I} + \Sigma \mathbf{A}_{ij}|^{1/2}} \times \exp \left[-\frac{1}{2} \{ \bar{\gamma}^2 \boldsymbol{\mu}_T' \Sigma^{-1} \boldsymbol{\mu}_T - (\bar{\gamma} \Sigma^{-1} \boldsymbol{\mu}_T + \mathbf{B}_{ij})' (\Sigma^{-1} + \mathbf{A}_{ij})^{-1} (\bar{\gamma} \Sigma^{-1} \boldsymbol{\mu}_T + \mathbf{B}_{ij}) \} \right]$$

and

$$(11) \quad q^{(k\ell)} = C \exp \left[-\frac{1}{2} (\Gamma_{k\ell} - \mathbf{A}_{ij}^{-1} \mathbf{B}_{ij})' \mathbf{A}_{ij} (\Gamma_{k\ell} - \mathbf{A}_{ij}^{-1} \mathbf{B}_{ij}) - \mathbf{B}_{ij}' \mathbf{A}_{ij}^{-1} \mathbf{B}_{ij} \right]$$

for $(k, \ell) \neq (i, j)$, with C chosen to satisfy $q_0^{(ij)} + \sum_{(k, \ell) \neq (i, j)} q^{(k\ell)} = 1$. Observe that unlike all DP-based approaches hitherto considered in the statistics literature, in our case $\mathbf{G}_{ij}^{(T)}$, the conditional posterior base measure is not independent of $\mathbf{\Gamma}_{i'j'}$ for $(i', j') \neq (i, j)$, which is a consequence of the fact that, thanks to (2), $\mathbf{\Gamma}_{i'j'}$ are not conditionally independent of each other. Thus, our methodology generalizes other DP-based methods, including that of Gelfand, Kottas and MacEachern (2005).

Section S-1.2 presents an alternative algorithm to updating $\mathbf{\Gamma}_{ij}$ using *configuration indicators* which are updated sequentially using themselves and only the distinct $\mathbf{\Gamma}_{ij}$, given everything else. MacEachern (1994) has argued that such an updating procedure theoretically improves convergence properties of the Markov chain: however, Section S-1.3 shows that in our case the associated conditional distributions need to be obtained separately for each of the 2^9 possible configuration indicators. This being infeasible, we recommend (9) for updating $\mathbf{\Gamma}_{ij}$. [We remark here that full conditionals are easily obtained using configuration indicators in the case of Gelfand, Kottas and MacEachern (2005), thanks to the relative simplicity of their spatial problem.] Also, (10) and (11) imply that as $\tau \rightarrow \infty$, the full conditional distribution (9) converges to $\mathbf{G}_{ij}^{(T)}$, which is actually the full conditional distribution of the entire T -dimensional parameter vector $\mathbf{\Gamma}_{ij}$ under the AR(1) model. In either case, we provide computationally efficient multivariate updates for our Gibbs updates: this makes our problem computationally tractable.

To obtain the full conditional of τ , define $m = \#\{(i, j); i, j = 1, 2, \dots, R\} = R^2$. Then, note that as in Escobar and West (1995), for a $\Gamma(a_\tau, b_\tau)$ prior on τ , the full conditional distribution of the latter, given the number (d) of distinct $\mathbf{\Gamma}_{ij}$ and another continuous random variable η , is a mixture of two Gamma distributions, specifically $\pi_\eta \Gamma(a_\tau + d, b_\tau - \log(\eta)) + (1 - \pi_\eta) \Gamma(a_\tau + d - 1, b_\tau - \log(\eta))$, where $\pi_\eta / (1 - \pi_\eta) = (a_\tau + d - 1) / (m(b_\tau - \log(\eta)))$. Also, the full conditional of η is $\beta(\tau + 1, m)$. Finally, the full conditional distributions of σ_δ^2 and ρ are not very standard and need careful derivation. Section S-1.4 describes a Gibbs sampling approach using configuration sets for updating σ_δ and ρ . For implementing this Gibbs step, one does not need to simulate the configuration indicators, as they can be determined after simulating the $\mathbf{\Gamma}_{ij}$'s using (9). Hence, this step is feasible. However, we failed to achieve sufficiently good convergence with this approach, and hence used a Metropolis–Hastings step. The acceptance ratio for the Metropolis–Hastings step is given by $[\mathbf{\Gamma}_{11}][\mathbf{\Gamma}_{12} | \mathbf{\Gamma}_{11}][\mathbf{\Gamma}_{13} | \mathbf{\Gamma}_{12}, \mathbf{\Gamma}_{11}] \cdots [\mathbf{\Gamma}_{33} | \mathbf{\Gamma}_{32}, \dots, \mathbf{\Gamma}_{11}]$, evaluated, respectively, at the new and the old values of the parameters (σ_δ^2, ρ) . In the above, $[\mathbf{\Gamma}_{11}] \sim \mathbf{G}_0^{(T)}$, and the other factors are Polya urn distributions, following easily from (8). Once again, note the use of multivariate updates in the MCMC steps, making our updating approach computationally feasible and easily implemented.

We conclude this section by noting that our model is structured to be identifiable. The priors of α_i , $\beta_i(t)$, $\gamma_{ij}(t)$ are all different and informative. Further, (2) shows that $\beta_i(t)$ is not permutation-invariant with respect to the indices of Γ_{ij} 's. Identifiability of our model is further supported by the results in this paper, which show all posteriors (based on MCMC) to be distinct and different. This is unlike the case of the usual Dirichlet process-based mixture models which are permutation-invariant, as in Escobar and West (1995), where the parameters have the same posterior due to nonidentifiability. We now investigate performance of our methodology.

3. Simulation studies. We performed a range of simulation experiments to investigate performance of our approach relative to its alternatives. Since there are 9 nonzero Γ_{ij} 's in our model, we followed the recipe provided in Section 2 and put a $\Gamma(0.8, 0.1)$ prior on the DP scale parameter τ . We investigated fitting \mathcal{M}_{DP} , \mathcal{M}_{AR} and \mathcal{M}_{RW} to the simulated data of Section 1.2.2 and also to data simulated from the \mathcal{M}_{RW} and \mathcal{M}_{AR} models, the latter with both $\rho = 0.5$ (clearly stationary model) and $\rho = 0.95$ (where the model is not so clearly distinguished from nonstationarity but more clearly distinguished than when $\rho = 0.999$). The Gibbs sampling procedure for model \mathcal{M}_{AR} in our simulations was very similar to that of the \mathcal{M}_{RW} detailed in Bhattacharya, Ho and Purkayastha (2006): we omit details. For all experiments in this paper and in the supplement, we discarded the first 10,000 MCMC iterations as burn-in and stored the following 20,000 iterations for Bayesian inference. Our results are summarized here for want of space, but presented in detail in Section S-2, with performance evaluated graphically [in terms of the posterior densities of $\gamma_{ij}(t)$'s] and numerically using coverage and average lengths of the 95% HPD credible intervals of the posterior predictive distributions (for details, see Section S-2).

The results of our experiments using the simulated data of Section 1.2.2 showed that \mathcal{M}_{AR} performed better than \mathcal{M}_{RW} , but model \mathcal{M}_{DP} was the clear winner. Indeed, the support of the posterior distributions of $\gamma_{22}(t)$ and $\gamma_{23}(t)$ using \mathcal{M}_{AR} were much too wide to be of much use, but substantially narrower under \mathcal{M}_{DP} . \mathcal{M}_{DP} also outperformed the other two models in terms of the proportion of true $\gamma_{ij}(t)$'s included in the corresponding 95% HPD CIs. These CIs also captured almost all of the true values of $\gamma_{ij}(t)$ under \mathcal{M}_{DP} , but far fewer values using \mathcal{M}_{AR} . \mathcal{M}_{DP} also exhibited better predictive performance than \mathcal{M}_{AR} and \mathcal{M}_{RW} . All these findings which favor our DP-based model were implicitly the consequence of the fact that the true model in our experiment was approximately nonstationary, and modeled more flexibly by our nonstationary DP model rather than the stationary AR(1) model. That this borderline between stationarity and nonstationarity of the true model is important was vindicated by the results of fitting \mathcal{M}_{RW} , \mathcal{M}_{AR} and \mathcal{M}_{DP} on the dataset simulated using \mathcal{M}_{RW} . Here, \mathcal{M}_{RW}

outperformed both \mathcal{M}_{DP} and \mathcal{M}_{AR} in terms of coverage of the true values of $\gamma_{ij}(t)$, indicating that \mathcal{M}_{DP} may under-perform when compared to the true model, in terms of coverage of parameter values, when the true model can be clearly identified. In terms of prediction ability, however, \mathcal{M}_{DP} was still the best performer, with the best coverage of the data points by the posterior predictive distribution and the lengths of the associated 95% CIs. This finding was not unexpected, since \mathcal{M}_{DP} involves model averaging (see Section S-1.5), which improves predictive performance [see, e.g., Kass and Raftery (1995)]. For the dataset simulated from \mathcal{M}_{AR} with $\rho = 0.5$, the true model (\mathcal{M}_{AR}) outperformed \mathcal{M}_{DP} marginally and \mathcal{M}_{RW} substantially, but when $\rho = 0.95$, \mathcal{M}_{DP} provided a much better fit than \mathcal{M}_{AR} or \mathcal{M}_{RW} . We have already mentioned that \mathcal{M}_{DP} outperformed \mathcal{M}_{AR} (and \mathcal{M}_{RW}) for the borderline case of $\rho = 0.999$: the experiment with $\rho = 0.95$ demonstrated good performance of \mathcal{M}_{DP} even in relatively more distinguishable situations. At the same time, the experiment with $\rho = 0.5$ warns against over-optimism regarding \mathcal{M}_{DP} ; for clearly stationary data, we are at least marginally better off replacing \mathcal{M}_{DP} with a stationary model such as \mathcal{M}_{AR} . In spite of this caveat for clearly stationary situations, our simulation experiments indicated that our DP-based approach is flexible enough to address stationary models as well as deviations. We now analyze the Stroop Task dataset introduced in Section 1.1.

4. Application to Stroop task data. The dataset was preprocessed following Ho, Ombao and Shumway (2005) and Bhattacharya, Ho and Purkayastha (2006), to which we refer for details while providing only a brief summary here. For each of the three regions (LG, MOG and DLPFC), a spherical region of 33 voxels was drawn around the location of peak activation. The voxel-wise time series of the selected voxels in each region were then subjected to higher order (multi-linear) singular value decomposition (HOSVD) using methods in Lathauwer, Moor and Vandewalle (2000). The first mode of this HOSVD, after detrending with a running-line smoother as in Marchini and Ripley (2000), provided us with our detrended time series response $y_i(t)$ for the i th region [see Figure S-4 for $y(t)$'s as well as $x(t)$].

We compared results obtained using \mathcal{M}_{DP} with those using \mathcal{M}_{RW} and \mathcal{M}_{AR} . We refer to Bhattacharya, Ho and Purkayastha (2006) and the supplement for detailed results using \mathcal{M}_{RW} and \mathcal{M}_{AR} , respectively, only summarizing them here in comparison with results obtained using \mathcal{M}_{DP} , which we also discuss in greater detail here. Detailed studies on MCMC convergence are in Section S-3.2.

4.1. Results. Figure 2 displays the Gibbs-estimated marginal posterior distributions of the $\gamma_{ij}(t)$'s for each time point t obtained using \mathcal{M}_{DP} . A striking feature of the marginal posterior densities of Figure 2 is the very

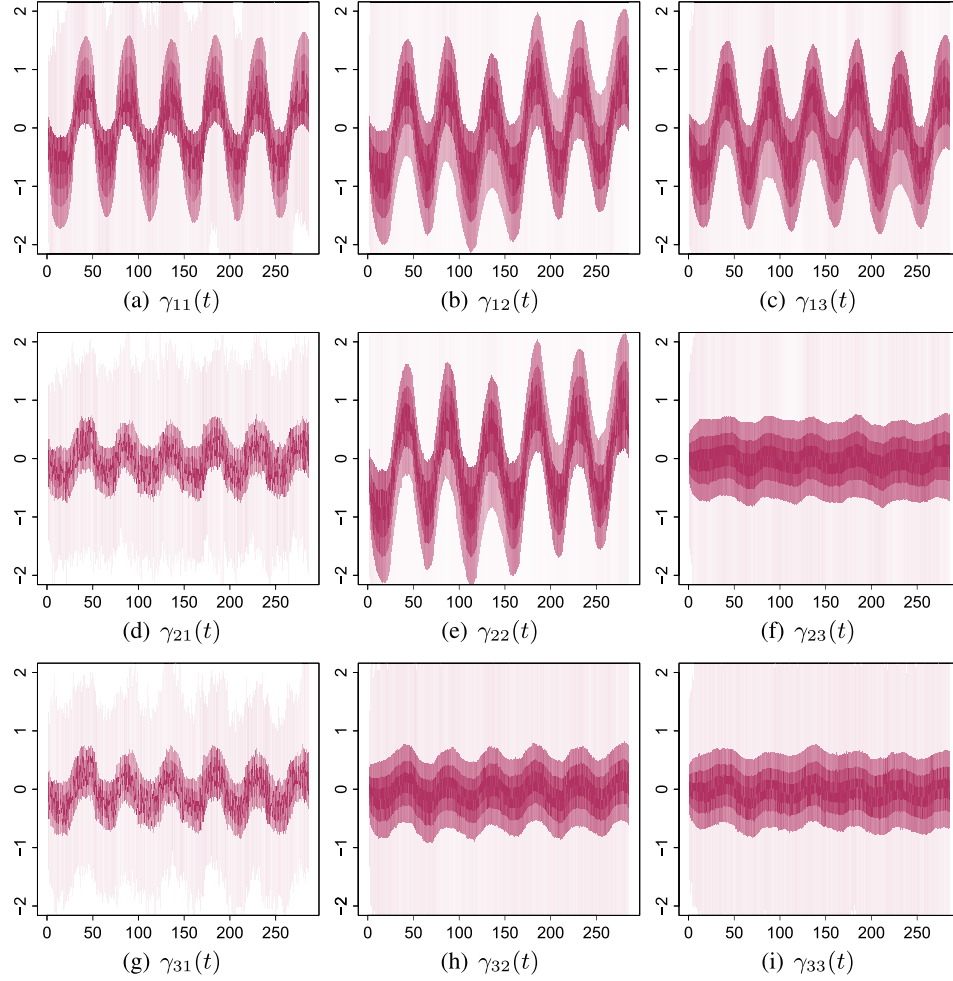


FIG. 2. *Estimated posterior densities (means in solid lines) of the regional influences over time.*

strong oscillatory nature of these effective connectivity parameters with the modeled BOLD response $x(t)$. This is quite different from the posterior distributions of $\gamma_{ij}(t)$'s obtained using \mathcal{M}_{AR} (see Figure S-7). Table 2 evaluates performance of the two models in terms of the length and proportion of observations contained in the 95% HPD credible intervals of the posterior predictive distributions: the intervals obtained using \mathcal{M}_{DP} have greater coverage but are also much narrower, making it by far the better choice among the models. Figure 2 also shows that $\gamma_{23}(t)$, $\gamma_{32}(t)$ and $\gamma_{33}(t)$ —and, to a lesser extent, $\gamma_{21}(t)$ and $\gamma_{31}(t)$ —oscillate differently from the others in that their amplitude is close to zero. We examined this issue further through

TABLE 2
Proportions of observed y included in, and average length, of the 95% credible intervals of the posterior predictive distributions under \mathcal{M}_{AR} and \mathcal{M}_{DP} for the Stroop task dataset

y	Proportions		Average length	
	\mathcal{M}_{AR}	\mathcal{M}_{DP}	\mathcal{M}_{AR}	\mathcal{M}_{DP}
y_1	0.92	0.99	4,960.9	2,215.1
y_2	1.00	1.00	3,864.2	2,068.1
y_3	1.00	1.00	4,352.8	2,084.3

Figure 3, which provides a map of the proportions of the cases for which each estimated marginal posterior density of $\gamma_{ij}(t)$ has positive support at time t . The intensities are mapped via a red-blue diverging palette: thus, darker hues of blue and red indicate high and low values, respectively, for the proportions. Lighter hues of red or blue indicate values in the middle. Clearly, very little proportion of the marginal density is either on the positive or the negative parts of the real line for the cases of $\gamma_{23}(t)$, $\gamma_{32}(t)$ and $\gamma_{33}(t)$. We therefore investigated performance of models \mathcal{M}_{DP} modified to exclude some or all of these regional influences.

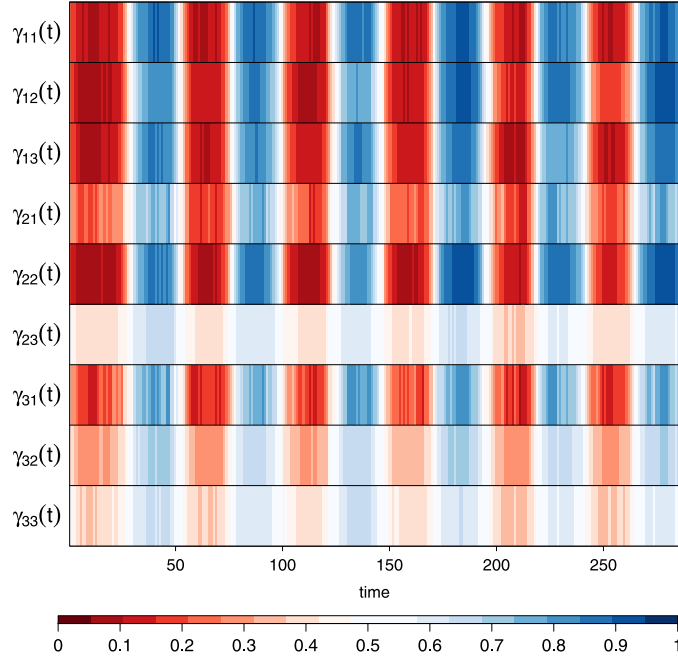


FIG. 3. *Proportions of estimated marginal posterior density of $\gamma_{ij}(t)$ with positive support at t .*

4.1.1. *Investigating restricted submodels of \mathcal{M}_{DP} .* Bhattacharya, Ho and Purkayastha (2006) found that the model \mathcal{M}_{RW} with the constraint $\gamma_{31}(t) = \gamma_{32}(t) = 0$ (henceforth $\mathcal{M}_{\text{RW}}^-$) provided better results than the unconstrained \mathcal{M}_{RW} . Figure 2 also points to the possibility that models with some $\gamma_{ij}(t) \equiv 0$ might provide better performance. We explored these aspects quantitatively using the models \mathcal{M}_{AR} and \mathcal{M}_{DP} , by considering the proportion of data contained in, and the average lengths of, the 95% HPD CIs of the corresponding posterior predictive distributions of $y_i(t); i = 1, 2, 3, t = 1, \dots, T$. A systematic evaluation of all possible submodels is computationally very time-consuming, so we investigated models with combinations of $\gamma_{31}(t) = \gamma_{32}(t) \equiv 0$ as in Bhattacharya, Ho and Purkayastha (2006) and with null $\gamma_{ij}(t)$'s for those (i, j) 's whose posterior distributions exhibited less amplitude of oscillation as per Figure 2. Table 3 summarizes performances of the top three submodels: others are in Tables S-11 and S-12. The top three performers were the following:

- $\mathcal{M}_{\text{DP}}^{(1)}$: \mathcal{M}_{DP} but with $\gamma_{33}(t) \equiv 0 \forall t$.
- $\mathcal{M}_{\text{DP}}^{(2)}$: \mathcal{M}_{DP} but with $\gamma_{32}(t) \equiv 0 \forall t$.
- $\mathcal{M}_{\text{DP}}^{(3)}$: \mathcal{M}_{DP} but with $\gamma_{32}(t) = \gamma_{33}(t) \equiv 0 \forall t$.

Thus, $\mathcal{M}_{\text{DP}}^{(1)}$ and $\mathcal{M}_{\text{DP}}^{(2)}$ both beat \mathcal{M}_{DP} (of Table 2). The average 95% posterior predictive length using $\mathcal{M}_{\text{DP}}^{(2)}$ is about midway between $\mathcal{M}_{\text{DP}}^{(1)}$ and the unrestricted DP-based model, so we report our final findings and conclusions only using $\mathcal{M}_{\text{DP}}^{(1)}$.

4.2. *Summary of findings.* Figure 4(a)–(h) display the posterior densities of the nonnull regional influences $\gamma_{ij}(t)$'s over time. These $\gamma_{ij}(t)$'s are very similar to those in Figure 2(a)–(h), with nonzero effective connectivity parameters again having a very pronounced oscillation synchronous with the modeled BOLD response: indeed, only the $\gamma_{23}(t)$ of Figure 4(f) has an

TABLE 3

Proportions of the observed data in, and mean lengths of, the 95% credible intervals of posterior predictive distributions of y_1, y_2, y_3 and the mean lengths of the 95% credible intervals for the top three candidate submodels

y	Proportion			Mean length		
	$\mathcal{M}_{\text{DP}}^{(1)}$	$\mathcal{M}_{\text{DP}}^{(2)}$	$\mathcal{M}_{\text{DP}}^{(3)}$	$\mathcal{M}_{\text{DP}}^{(1)}$	$\mathcal{M}_{\text{DP}}^{(2)}$	$\mathcal{M}_{\text{DP}}^{(3)}$
y_1	0.99	0.99	1.0	2,097.6	2,140.4	2,276.5
y_2	1.0	1.0	1.0	1,971.6	2,019.5	2,127.8
y_3	1.0	1.0	1.0	1,985.0	2,021.3	2,125.4

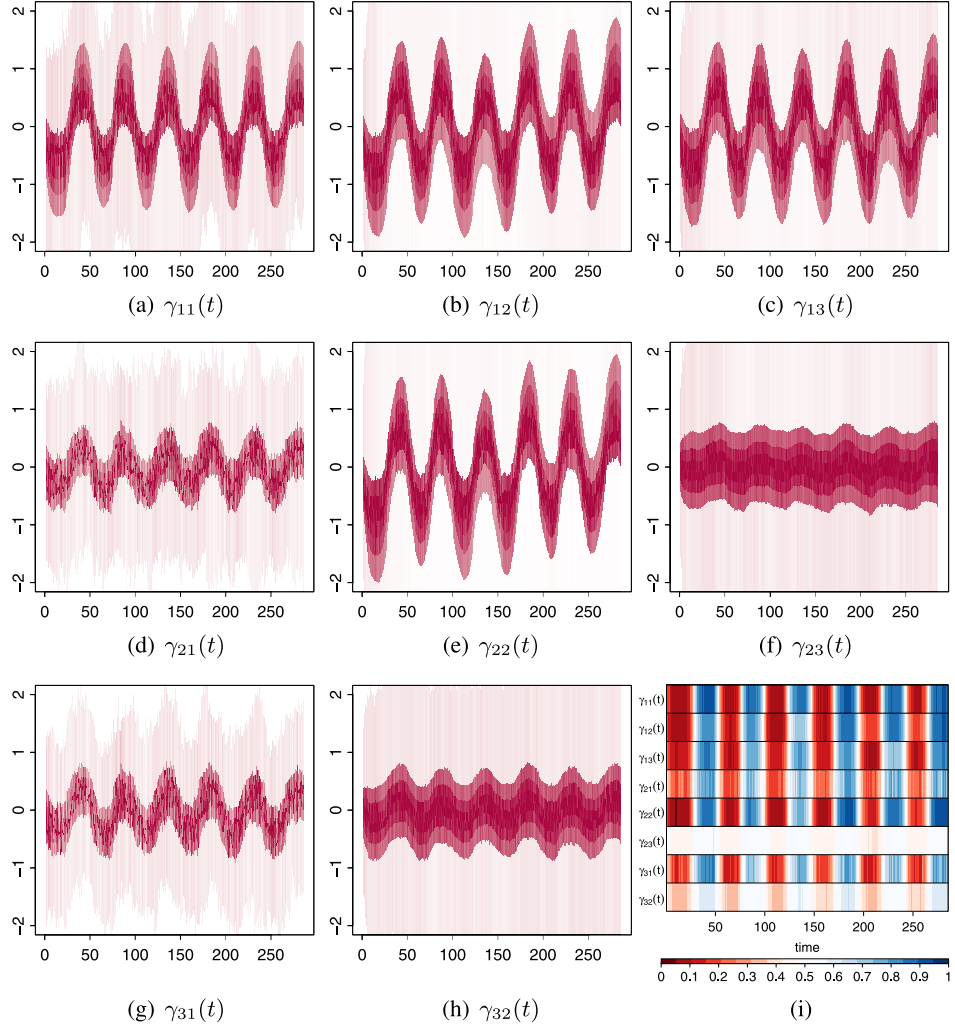


FIG. 4. (a)–(h) *Estimated posterior densities (means in solid lines) of the nonnull regional influences over time using $\mathcal{M}_{\text{DP}}^{(1)}$.* (i) *Proportion of the posterior distribution of $\gamma_{ij}(t)$ with positive support at time t .*

oscillation slightly more damped than in Figure 2. Further, Figure 4(i) indicates that the estimated posterior densities put most of their mass either below zero [when $x(t)$ is negative] or above zero [when $x(t)$ is positive]. Indeed, these densities have substantial mass around zero only when $x(t)$ is around zero. We also smoothed the modeled BOLD response $x(t)$ to explore further its relationship with each of the estimated posterior mean $\gamma_{ij}(t)$'s from $\mathcal{M}_{\text{DP}}^{(1)}$. For each t , we specified $x(t) = A \cos(2\pi\omega t + \phi) + \psi_t$ where ψ_t are i.i.d. $N(0, \sigma_\psi^2)$, A is the amplitude of the time series, ω is the oscilla-

tion frequency and ϕ is a phase shift. Equivalently, $x(t) = \beta_1 \cos(2\pi\omega t) + \beta_2 \sin(2\pi\omega t) + \psi_t$ with $\beta_1 = A \cos(\phi)$ and $\beta_2 = A \sin(\phi)$. We obtain $\hat{\omega} = 0.02$ using the periodogram approach [see, i.e., Shumway and Stoffer (2006)]. Thus, each cycle in $x(t)$ has a length of about 50 time-points. A least squares fit yields $\hat{\beta}_1 = 0.27$ and $\hat{\beta}_2 = -0.61$, hence, $\hat{A} = 0.80$ and $\phi = 1.16$. Figure S-8 shows that the smoothed BOLD response $\hat{x}(t) = \hat{\beta}_1 \cos(2\pi\hat{\omega}t) + \hat{\beta}_2 \sin(2\pi\hat{\omega}t)$ closely approximates the original time series $x(t)$. The correlation of $\hat{x}(t)$ with each of $\gamma_{11}(t)$, $\gamma_{12}(t)$, $\gamma_{13}(t)$, $\gamma_{21}(t)$, $\gamma_{22}(t)$, $\gamma_{23}(t)$, $\gamma_{31}(t)$ and $\gamma_{32}(t)$ are 0.959, 0.909, 0.952, 0.950, 0.922, 0.874, 0.949 and 0.929, respectively. Thus, $\gamma_{ij}(t)$'s are not completely linear in the BOLD response, but very close to being so with regard to its transformed version.

The results of our analysis indicate that the region LG, centered around zero, exhibits very strong evidence of self-feedback, oscillatory with high amplitude, and period of about 50, matching the period of the modeled BOLD response $x(t)$. Similar influences are exerted by both MOG and DLPFC on LG and by the MOG region on itself. Indeed, Figure 4 indicates that these four inter- and intra-regional influences have, broadly, a similar pattern in terms of amplitude. The influence of LG on MOG and DLPFC is smaller and similar to each other. Further, Figure 4(f) and (h) indicate that the feedback provided by DLPFC on MOG [$\gamma_{23}(t)$] is similar to that in the reverse direction [$\gamma_{32}(t)$]. Thus, there are three broad patterns in the way that inter- and intra-regional influences occur.

Our analysis also demonstrates the existence of the ACN and its mechanism while performing a Stroop task. Thus, the executive control system (DLPFC) provides instruction to both the task-irrelevant (LG) and task-relevant processing sites (MOG) but gets similar levels of feedback from the task-relevant processor (MOG). LG which sifts out the task-irrelevant color information gets a lot of feedback in doing so from both itself and MOG. However, it provides far less feedback to the task-relevant shape information processing MOG and the executive control DLPFC. MOG itself provides substantial self-feedback while processing shape information. Finally, note that while our results indicate higher amplitudes for inter-regional feedback involving $\gamma_{ij}(t)$'s when they involve LG rather than MOG, this is consistent with the established notion that processing shape information is a higher-level (more difficult) cognitive function than distinguishing color.

The results on the effective connectivity parameters using \mathcal{M}_{DP} are very different from those done using \mathcal{M}_{RW}^- [see Figure 5 of Bhattacharya, Ho and Purkayastha (2006)] or \mathcal{M}_{AR} . Using \mathcal{M}_{RW}^- , Bhattacharya, Ho and Purkayastha (2006) found some evidence of self-feedback only in LG: the 95% HPD BCRs contained zero unless t increased. Further, while the relationship of the posterior mean appeared somewhat linear in t , there was no relationship with the modeled BOLD response. Most $\gamma_{ij}(t)$'s [with the

exception of $\gamma_{13}(t)$] were almost invariant with respect to time t , unlike the clear oscillatory nature of the time series obtained here using $\mathcal{M}_{\text{DP}}^{(1)}$ (or even \mathcal{M}_{DP}). The fact that the BOLD response had very little relationship with these effective connectivity parameters is perplexing, given that these regions were the ones found to be activated in the preprocessing of the fMRI dataset. The results on $\gamma_{ij}(t)$'s using \mathcal{M}_{AR} were also very surprising: while the posterior means oscillated synchronously with $x(t)$ only for the task-irrelevant LG with a correlation of 0.943, there was no evidence of nonzero values for all the other effective connectivity parameter values (including the task-relevant MOG), since their pointwise 95% HPD credible regions all contained zero for all time t . This is very unlike the results obtained using $\mathcal{M}_{\text{DP}}^{(1)}$, which also established the existence of the ACN theory in performing this task. Indeed, among all the approaches considered in the literature and here on this dataset, only the DP-based analyses have been able to capture both the dynamic as well as the oscillatory nature of the effective connectivity parameters. In doing so, we also obtain further insight into how an individual brain performs a Stroop task.

5. Conclusions and future work. Effective connectivity analysis provides an important approach to understanding the functional organization of the human brain. Bhattacharya, Ho and Purkayastha (2006) provide a coherent and elegant Bayesian approach to incorporating uncertainty in the analysis. In this paper we note that this approach also brings forth with it some limitations. In this paper we therefore propose a nonstationary and nonparametric Bayesian approach using a DP-based model that embeds an AR(1) process in the class of many possible models. Heuristically, our suggestion has some connection with model averaging, where we have, a priori, an AR(1) model in mind for specifying dynamic effective connectivity: the DP provides a coherent way to formalize our intuition. We have also derived an easily implemented Gibbs sampling algorithm for learning about the posterior distributions of all the unknown quantities. Simulation studies show that our model is a better candidate for the analysis of effective connectivity in many cases. The advantage is more pronounced with increasing departures from stationarity in the true model. We also applied our methodology to investigate the feedback mechanisms between the task-irrelevant LG, the task-relevant MOG and the “executive control” DLPFC in the context of a single-subject Stroop task study. Our results showed strong self-feedback for LG and MOG, but not for DLPFC. Further, MOG and DLPFC influence LG strongly but the reverse is rather mild. The influence of MOG on DLPFC and vice versa are very similar. All these discovered feedback mechanisms oscillate strongly in the manner of the BOLD signal and are supportive of the framework postulated by ACN theory. Our analysis also provides understanding into the mechanism of how the brain performs a Stroop task. All

these are novel findings not reported in the context of fMRI analysis in the literature. Thus, adoption of our DP-based approach not only provided interpretable results, but—as very kindly pointed out by a reviewer—yielded additional insights into the workings of the brain.

There are several aspects of our methodology and analysis that deserve further attention. For one, we have investigated ACN in the context of a Stroop task for a single male volunteer. It would be of interest to study other tasks and responses to other stimuli and also to see how our results on a Stroop task translate to multiple subjects and to investigate how these mechanisms differ from one person to another. Our modeling approach can easily be extended to incorporate such scenarios. Further, our methodology, while developed and evaluated in the context of modeling dynamic effective connectivity in fMRI datasets, can be applied to other settings also, especially in situations where the actual models for the unknowns may be quite difficult to specify correctly. Thus, we note that while this paper has made an interesting contribution to analyzing dynamic effective connectivity in single-subject fMRI datasets, several interesting questions and extensions meriting further attention remain.

SUPPLEMENTARY MATERIAL

Contents (DOI: [10.1214/11-AOAS470SUPP](https://doi.org/10.1214/11-AOAS470SUPP); .pdf). Section S-1 contains additional details regarding our methodology, including explicit forms of the full conditional distributions of specific parameters, the configuration indicators and the distinct parameters associated with the Dirichlet process needed for Gibbs sampling. Detailed arguments that show model averaging associated with our DP-based model \mathcal{M}_{DP} are also presented there. Section S-2 provides additional information on our simulation experiments, including associated methodology and results. Section S-3 presents further details on the analysis of the Stroop task experiment, including display of the data, detailed assessment of convergence of our MCMC samplers when using \mathcal{M}_{DP} and $\mathcal{M}_{DP}^{(1)}$ and MCMC-based posterior analysis using \mathcal{M}_{AR} and other additional models obtained by setting some effective connectivity parameters to zero. Additional methodological details and results regarding the smoothing of the modeled BOLD signal $x(\cdot)$ are also presented there.

Acknowledgments. The authors are very grateful to the Editor and two reviewers, whose very detailed and insightful comments on earlier versions of this manuscript greatly improved its content and presentation.

REFERENCES

- AERTSEN, A. and PREIBL, H. (1991). Dynamics of activity and connectivity in physiological neuronal networks. In *Non-linear Dynamics and Neuronal Networks* (H. G. SCHUSTER, ed.) 281–302. VCH, New York.

- BANACH, M. T., MILHAM, M. P., ATCHLEY, R., COHEN, N. J., WEBB, A., WSZALEK, T., KRAMER, A. F., LIANG, Z. P., WRIGHT, A., SHENKER, J. and MAGIN, R. (2000). fMRI studies of Stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. *Journal of Cognitive Neuroscience* **12** 988–1000.
- BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer, New York. [MR0804611](#)
- BHATTACHARYA, S., HO, M. R. and PURKAYASTHA, S. (2006). A Bayesian approach to modeling dynamic effective connectivity with fMRI data. *NeuroImage* **30** 794–812.
- BHATTACHARYA, S. and MAITRA, R. (2011). Supplement to “A nonstationary nonparametric Bayesian approach to dynamically modeling effective connectivity in functional magnetic resonance imaging experiments.” [DOI:10.1214/11-AOAS470SUPP](#).
- BÜCHEL, C. and FRISTON, K. J. (1998). Dynamic changes in effective connectivity characterized by variable parameter regression and Kalman filtering. *Human Brain Mapping* **6** 403–408.
- BUXTON, R. B., WONG, E. C. and FRANK, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magnetic Resonance in Medicine* **39** 855–864.
- CORBETTA, M., MIEZIN, F. M., DOBMEYER, S., SHULMAN, G. L. and PETERSEN, S. E. (1991). Selective and divided attention during visual discrimination of shape, color and speed: Functional anatomy by positron emission tomography. *Journal of Neuroscience* **8** 2383–2402.
- DE IORIO, M., MÜLLER, P., ROSNER, G. L. and MACEACHERN, S. N. (2004). An ANOVA model for dependent random measures. *J. Amer. Statist. Assoc.* **99** 205–215. [MR2054299](#)
- ESCOBAR, M. D. and WEST, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90** 577–588. [MR1340510](#)
- FRISTON, K. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping* **2** 56–78.
- FRISTON, K. J. (2011). Dynamic causal modeling and Granger causality Comments on: The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. *NeuroImage*. To appear.
- FRISTON, K. J., HARRISON, L. and PENNY, W. (2003). Dynamic causal modeling. *Neuroimage* **19** 1273–1302.
- FRISTON, K. J., MECHELLI, A., TURNER, R. and PRICE, C. J. (2000). Nonlinear responses in fMRI: The Balloon model, Volterra kernels, and other hemodynamics. *Neuroimage* **12** 466–477.
- FRITH, C. (2001). A framework for studying the neural basis of attention. *Neuropsychologia* **39** 167–1371.
- GELFAND, A. E., KOTTAS, A. and MACEACHERN, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *J. Amer. Statist. Assoc.* **100** 1021–1035. [MR2201028](#)
- GLOVER, G. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage* **9** 416–429.
- GOEBEL, R., ROEBROECK, A., KIM, D. S. and FORMISANO, E. (2003). Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magnetic Resonance Imaging* **21** 1251–1261.
- GÖSSL, C., AUER, D. P. and FAHRMEIR, L. (2001). Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics* **57** 554–562. [MR1855691](#)
- HARRISON, L., PENNY, W. L. and FRISTON, K. (2003). Multivariate autoregressive modeling of fMRI time series. *Neuroimage* **19** 1477–1491.

- HARRISON, L., STEPHAN, K. E. and FRISTON, K. J. (2007). Effective connectivity. In *Statistical Parametric Mapping: The Analysis of Functional Brain Images* 508–521. Academic Press, New York.
- HENSON, R. and FRISTON, K. J. (2007). Convolution models for fMRI. In *Statistical Parametric Mapping: The Analysis of Functional Brain Images* 178–192. Academic Press, New York.
- HO, M. R., OMBAO, H. and SHUMWAY, R. (2003). Practice-related effects demonstrate complementary role of anterior cingulate and prefrontal cortices in attentional control. *NeuroImage* **18** 483–493.
- HO, M. R., OMBAO, H. and SHUMWAY, R. (2005). A state-space approach to modelling brain dynamics. *Statist. Sinica* **15** 407–425. [MR2190212](#)
- JAENSCH, E. R. (1929). *Grundformen Menschlichen Seins*. Otto Elsner, Berlin.
- KASS, R. E. and RAFTERY, R. E. (1995). Bayes factors. *J. Amer. Statist. Assoc.* **90** 773–795.
- KELLEY, W. M., MIEZIN, F. M., McDERMOTT, K. B., BUCKNER, R. L., RAICHLE, M. E., COHEN, N. J., OLLINGER, J. M., AKBUDAK, E., CONTURO, T. E., SNYDER, A. Z. and PETERSEN, S. E. (1998). Hemispheric specialization in human dorsal frontal cortex and medial temporal lobe for verbal and nonverbal memory encoding. *Neuron* **20** 927–936.
- KIRK, E., HO, M. R., COLCOMBE, S. J. and KRAMER, A. F. (2005). A structural equation modeling analysis of attentional control: An event-related fMRI study. *Cognitive Brain Research* **22** 349–357.
- LATHAUWER, L. D., MOOR, B. D. and VANDEWALLE, J. (2000). A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* **21** 1253–1278. [MR1780272](#)
- LINDQUIST, M. A. (2008). The statistical analysis of fMRI data. *Statist. Sci.* **23** 439–464. [MR2530545](#)
- LU, Y., BAGSHAW, A. P., GROVA, C., KOBAYASHI, E., DUBEAU, F. and GOTMAN, J. (2006). Using voxel-specific hemodynamic response function in EEG-fMRI data analysis. *Neuroimage* **32** 238–247.
- LU, Y., BAGSHAW, A. P., GROVA, C., KOBAYASHI, E., DUBEAU, F. and GOTMAN, J. (2007). Using voxel-specific hemodynamic response function in EEG-fMRI data analysis: An estimation and detection model. *Neuroimage* **34** 195–203.
- MACEachern, S. N. (1994). Estimating normal means with a conjugate-style Dirichlet process prior. *Comm. Statist. Simulation Comput.* **23** 727–741. [MR1293996](#)
- MACEachern, S. N. (2000). Dependent Dirichlet processes. Technical report, Dept. Statistics, Ohio State Univ., Columbus, OH.
- MARCHINI, J. L. and RIPLEY, B. D. (2000). A new statistical approach to detecting significant activation in functional MRI. *NeuroImage* **12** 366–380.
- McINTOSH, A. R. (2000). Towards a network theory of cognition. *Neural Networks* **13** 861–870.
- McINTOSH, A. R. and GONZALEZ-LIMA, F. (1994). Structural equation modeling and its application to network analysis of functional brain imaging. *Human Brain Mapping* **2** 2–22.
- MILHAM, M. P., BANICH, M. T. and BARAD, V. (2003). Competition for priority in processing increases prefrontal cortex’s involvement in top-down control: An event-related fMRI study of the Stroop task. *Cognitive Brain Research* **17** 212–222.
- MILHAM, M. P., ERICKSON, K. I., BANICH, M. T., KRAMER, A. F., WEBB, A., WSZALEK, T. and COHEN, N. J. (2002). Attentional control in the aging brain: Insights from an fMRI study of the Stroop task. *Brain Cognition* **49** 277–296.

- MILHAM, M. P., BANICH, M. T., CLAUS, E. and COHEN, N. (2003). Practice-related effects demonstrate complementary role of anterior cingulate and prefrontal cortices in attentional control. *Neuroimage* **18** 483–493.
- NYBERG, L. and MCINTOSH, A. R. (2001). Functional neuroimaging: Network analysis. In *Handbook of Functional Neuroimaging of Cognition* (R. CABEZA and A. KINGSTONE, eds.) 49–72. MIT Press, Cambridge, MA.
- PATRIOTA, A. G., SATO, J. R. and ACHIC, B. G. B. (2010). Vector autoregressive models with measurement errors for testing Granger causality. *Stat. Methodol.* **7** 478–497.
- PENNY, W. D., STEPHAN, K. E., MECHELLI, A. and FRISTON, K. J. (2004). Modeling functional integration: A comparison of structural equation and dynamic causal models. *Neuroimage* **23** (Suppl. 1) 264–274.
- RYKHEVSKAIA, E., FABIANI, M. and GRATTON, G. (2006). Lagged covariance structure models for studying functional connectivity in the brain. *Neuroimage* **30** 1203–1218.
- SATO, J. R., MORRETTIN, P. A., ARANTES, P. R. and AMARO JR., E. (2007). Wavelet-based time-varying vector autoregressive modeling. *Neuroimage* **51** 5847–5866. [MR2407682](#)
- SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4** 639–650. [MR1309433](#)
- SHUMWAY, R. H. and STOFFER, D. S. (2006). *Time Series Analysis and Its Applications With R Examples*. Springer, New York. [MR2228626](#)
- STEPHAN, K. E., WEISKOPF, N., DRYSDALE, P. M., ROBINSON, P. A. and FRISTON, K. J. (2007). Comparing hemodynamic models with DCM. *Neuroimage* **38** 387–401.
- STROOP, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* **18** 643–662.
- THOMPSON, W. K. and SIEGLE, G. (2009). A stimulus-locked vector autoregressive model. *Neuroimage* **46** 739–748.
- WORSLEY, K. J., LIAO, C. H., ASTON, J., PETRE, V., DUNCAN, G. H., MORALES, F. and EVANS, A. C. (2002). A general statistical analysis for fMRI data. *Neuroimage* **15** 1–15.

BAYESIAN AND INTERDISCIPLINARY
RESEARCH UNIT
INDIAN STATISTICAL INSTITUTE
203, B. T. ROAD
KOLKATA 700108
INDIA
E-MAIL: sourabh@isical.ac.in

DEPARTMENT OF STATISTICS
AND STATISTICAL LABORATORY
IOWA STATE UNIVERSITY
AMES, IOWA 50011-1210
USA
E-MAIL: maitra@iastate.edu